





AD-A170 110

			nEPORT DOCUME	NTATION PAGE	•		t				
1a REPORT	T SECURITY C	LASSIFICATION		1b. RESTRICTIVE MARKINGS							
	TY CLASSIFIC	ATION AUTHORITY		3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for Public Release; Distribution							
NA NA											
NA	SSIFICATION/	DOWNGRADING SCHE	·	Unlimited							
	MING ORGAN	IZATION REPORT NUR	ABER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 86-0382							
64 NAME C	OF PERFORMI	NG ORGANIZATION	SL OFFICE SYMBOL	7& NAME OF MONITORING ORGANIZATION							
Georgia	Institut	e of Technology	(If applicable)	AFOSR/NM							
Sc. ADDRE	SS (City, State	end ZIP Code)		7b. ADDRESS (City, State and ZIP Code)							
Atlan	ta, GA 30	332		Bldg. 410 Bolling AFB, DC 20332-6448							
	OF FUNDING/S	SPONSORING	6b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR 84-0367							
		4 7/B C - 4 - 1	<u> </u>								
Bldg	188 (City, State 3. 410		!	10. SOURCE OF FUI PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT				
Boll	ing AFB,	DC		6.1102F	2304	0.5					
spectio	Include Securion Periods	tor M/G/1 Que	mal Idle and In- ues			A.5					
	OF REPORT		COVERED	14. DATE OF REPORT (Yr., Mo., Dey)							
	orint EMENTARY N	FROM	<u> </u>	March 25,19	986		12				
10. 0077 2.		012110N									
17.	COSATI		18. SUBJECT TERMS (C	ontinue on reverse if n	ecessery and identi	ly by block numb	(P)				
FIELD	GROUP	SUB. GR.	-								
We consider an M/G/1 queue that operates under a (T,N)-policy: whenever the system becomes empty, the server is idle for a time T and then it inspects the queue continuously without serving customers until there are N customers waiting - thereupon the server is activated for service and serves customers continuously until the system becomes empty. This idle-inspection-service cycle is repeated indefinitely. There are costs for inspecting the queue, activating and running the server; and holding customers in the system. We present a computational procedure for determining the design parameters (T,N) that minimizes the average cost. 21. ABSTRACT SECURITY CLASS TION UNCLASSIFIED UNCLASSIFIED											
UNCLASS	IFIED/UNLIMI	TED SAME AS RP	T. DTIC USERS	UNCLASSIF	TED	U	·				
	e of mespons chard F.	Serfozo		(Include Ares C	2b. TELEPHONE NUMBER (Include Area Code) (404) 894-2305						

OPTIMAL IDLE AND INSPECTION PERIODS FOR M/G/1 QUEUES

bу

Sung Shick Kim Korea University

and

Richard F. Serfozo Georgia Institute of Technology

Abstract

We consider an M/G/1 queue that operates under a (T,N)-policy: whenever the system becomes empty, the server is idle for a time T and then it inspects the queue continuously without serving customers until there are N customers waiting - thereupon the server is activated for service and serves customers continuously until the system becomes empty. This idle-inspection-service cycle is repeated indefinitely. There are costs for inspecting the queue, activating and running the server, and holding customers in the system. We present a computational procedure for determining the design parameters (T,N) that minimize the average cost.

This research was supported in part by the grant AFOSR 84-0367.

Parameter description interests forester described

Approved for public release; distribution unlimited.

Introduction

Intermittent rather than continuous service is characteristic of service systems in which servers must be absent periodically for other duties or for rejuvenation. Intermittent service is also used in systems where short queues are tolerable, or where short busy periods for servers are uneconomical. In designing such systems, a natural question is: How long should the server be absent without observing the queue and at what queue length should the server start serving customers?

In this paper, we address this question for an M/G/1 queue that operates under a (T,N)-policy as follows. Customers arrive by a Poisson process $\{A(t); t > 0\}$ with rate λ , and the service times have a mean μ^{-1} and a finite variance, and $\mu > \lambda$. For simplicity, we assume the system begins at time zero with the server deactivated and no customers in the queue. The server remains idle in the time interval $\{0,T\}$ and, at time T, the queue is inspected which reveals A(T) customers waiting. If A(T) is less than a number N, then the queue is inspected continuously until the time S_N of the N-th arrival. Thereupon the server is activated for service and serves customers continuously until the system becomes empty, at which time the server is deactivated. On the other hand, if at time T it is found that A(T) > N, then the server is immediately activated for service and serves customers until the system becomes empty, as in the previous case. This idle-inspection-service cycle is repeated indefinitely.

Associated with this (T,N)-policy are costs for inspecting the queue, for activating and running the server, and for holding customers in the system. The aim is to find the design parameters (T,N) that minimize the average cost of operating the system.



A-1

:es

We begin our analysis in Section 1 by deriving an expression for the average cost of a (T,N)-policy. Then in Section 2 we present a method for computing an optimal (T,N)-policy. We also give some insights into how the optimal policy changes as the parameters of the model change.

The special (T,N)-policy with T=0 (no idle time) is the well-known N-policy studied by Yadin and Naor (1963), Heyman (1968), and Sobel (1969); related works are Bell (1971), Balachandran (1973), Levy and Yechiali (1975), Tijms (1976), Talman (1979), Shanthikumar (1981), Kimura (1982), and Lu and Serfozo (1984). Also, the special (T,N)-policy with N=1 (no inspection period) is essentially the T-policy studied in Heyman (1977) (Heyman and Sobel (1982) discuss the N- and T-policy as well). In He,man's model, when the server completes an idle period and finds no customers waiting, then the server takes another idle period; in our model the server is committed to serve after each idle period. We show how our analysis can be easily modified to conform to the former assumption. In addition to finding an optimal (T,N)-policy, our model is useful for comparing the costs of various (T,N)-policies in which one parameter is fixed and the other one is optimized.

I. The Average Cost of a (T,N)-Policy

In this section, we derive an expression for the average cost of operating the M/G/I queue under a fixed (T,N)-policy. We begin by introducing more notation.

Associated with the idle-inspection-service cycle described above, we let ι denote the length of time that the queue is inspected after time T, while no services are being performed. Namely, $\iota = \max\{0, S_N - T\}$. At time $T + \iota$ the server begins a busy period. The number of customers

waiting at the start of this busy period is $v = \max\{A(T), N\}$. We let B_v denote the length of the busy period starting with v customers. Then the total duration of the idle-inspection-service cycle is $S = T + \iota + B_v$.

We assume that the costs of operating the system are as follows:

K = cost per cycle for activating and deactivating the server

v = cost per unit time of inspecting (viewing) the queue

r = cost per unit time of running the server

h = cost per unit time of holding one customer in the system. Then the total cost for a cycle is

$$Z = K + v_1 + rB_v + h \int_0^Z X(t)dt,$$

where $\{X(t); t > 0\}$ is the number of customers in the system over time. The integral is the total customer waiting time.

Our main concern is the average cost per unit time over the infinite horizon, which we denote by C(T,N). Since the traffic intensity $\rho=\lambda/\mu$ is below one, the queueing process is regenerative, and so it is well known that C(T,N)=EZ/ES.

An expression for this cost is given in the following result. For this, we let τ denote the length of a busy period for a standard M/G/1 queue started with one customer, and let W denote the total waiting time of the customers present in this busy period. It is known (e.g. see p. 447 in Heyman and Sobel (1982)), that

$$(1.1) E\tau = 1/(u - \lambda)$$

(1.2)
$$EW = \lambda \sigma^2 / (2(1 - \rho)^2) + 1/(\mu - \lambda),$$

where σ^2 is the variance of the service time. We also let

$$\phi_1(T,N) = \sum_{n=0}^{N-1} (N-n)(\lambda T)^n e^{-\lambda T}/n!$$

$$\phi_2(T,N) = \sum_{n=0}^{N-1} (N^2 - n^2)(\lambda T)^n e^{-\lambda T}/n.$$

Theorem. Under the preceding assumptions,

(1.3) ES =
$$[T+\lambda^{-1}\phi_1(T,N)]/(1-\rho)$$

(1.4) EZ = K + [h(
$$\lambda$$
T)² + h ϕ_2 (T,N) + ϕ_1 (T,N)(2v(1- ρ) - h)]/[2 λ (1- ρ)]

(1.5)
$$C(T,N) = (h/2) \{ [a\lambda K + (\lambda T)^2 + (av-1)\phi_1(T,N) + \phi_2(T,N)] / [\lambda T + \phi_1(T,N)] \} + r\rho + h\lambda(1-\rho)EW,$$

where $a = 2(1-\rho)/h$.

Proof. By the definition of the cycle time S, we have

(1.6)
$$ES = T + E_1 + E_{1}$$

Clearly

(1.7)
$$E_{1} = E[E(\max\{0, S_{N} - T\} | A(T))]$$

$$= \sum_{n=0}^{N-1} E(S_{N-n}) P(A(T) = n) = \phi_{1}(T, N) / \lambda.$$

Next, we can write $B_{\nu} = \sum_{k=1}^{\nu} \tau_k$, where τ_1, τ_2, \ldots are independent copies of τ that are independent of ν . Then by a standard conditioning argument and (1.1), we have

(1.8)
$$EB_{\nu} = E\nu E\tau = E\nu/(\mu - \lambda),$$

where

(1.9)
$$E_{\nu} = E[A(T) + \max\{0, N - A(T)\}]$$
$$= \lambda T + \phi_{1}(T, N).$$

Combining (1.6) - (1.9) yields expression (1.3).

Now consider the expected cycle cost

(1.10)
$$EZ = K + vE_1 + rEB_v + hE \int_0^Z X(t)dt.$$

We already have expressions for E1 and EB $_{\nu}$. It remains to find an expression for the expectation of the waiting time

By Fubini's theorem, we have

(1.12)
$$E \int_0^T A(t)dt = \int_0^T EA(t)dt = \int_0^T \lambda t dt = \lambda T^2/2.$$

Next, observe that

(1.13)
$$E \int_{T}^{T+1} A(t)dt = E[I(A(T) < N) \sum_{n=A(T)}^{N-1} nY_{n}]$$

$$= \lambda^{-1} E[I(A(T) < N) \sum_{n=A(T)}^{N-1} n],$$

$$= nA(T)$$

where Y_1 , Y_2 ,... are independent exponential variables with mean λ^{-1} that are independent of A(T), and I is the indicator function. Then applying the identity

$$N-1$$
 Σ
 $n = [(N-1)N - m(m+1)]/2 = [(N^2-m^2) - (N-m)]/2$
 $n=m$

to (1.13), and recalling the definitions of ϕ_1 and ϕ_2 , we obtain

(1.14)
$$E \int_{T}^{T+t} A(t)dt = \left[\phi_{2}(T,N) - \phi_{1}(T,N)\right]/(2\lambda).$$

Finally, we can write

$$\int_{T+1}^{z} X(t)dt = \sum_{n=1}^{v} \int_{t}^{t} \frac{1}{v-n} X(t)dt,$$

where $t_n = \inf\{t > T + \iota$: X(t) = n }. In the time interval $[t_{v-n+1}, t_{v-n}]$, the process X(t) starts at v-n+1 and remains above this level until the end when it reaches v-n. Let (τ_1, W_1) , (τ_2, W_2) ,... denote independent copies of (τ, W) that are independent of v. Then

$$\int_{t}^{t} \frac{v-n}{v-n+1} X(t) dt = W_n + (v-n)t_n,$$

where W_n is the area under X(t) above the level v-n, and $(v-n)\tau_n$ is the area under X(t) below v-n ($t_{v-n} = T + \iota + \tau_1 + \cdots + \tau_n$). It follows that

(1.15)
$$E \int_{T+1}^{Z} X(t) dt = E \sum_{k=1}^{V} W_{k} + E \{ E \left[\sum_{k=1}^{V} (v-k) \tau_{k} \middle| v \right] \}$$

$$= E vEW + E \tau E \left[\sum_{k=1}^{V} (v-k) \right]$$

$$= E vEW + E \tau E (v^{2} - v) / 2,$$

where

(1.16)
$$Ev^{2} = E[A(T)^{2} + \max\{0, N^{2} - A(T)^{2}\}]$$
$$= \lambda T + (\lambda T)^{2} + \phi_{2}(T, N).$$

Substituting (1.11) - (1.16) into (1.10) yields expression (1.4). Then expression (1.5) follows from (1.3), (1.4) and C(T,N) = EZ/ES.

Computation of Optimal (T,N)-Policies

In this section, we address the problem of finding a (T,N)-policy that minimizes the average cost C(T,N).

As a first step, consider the subproblem of minimizing C(T,N) over N for T fixed. This is of interest in itself when one is designing a system in which the idle time T is preset and cannot be varied. The solution to this subproblem is as follows.

Theorem 2.1. For each T, the cost C(T,N) has a unique minimum over N, which is attained at the value

$$N(T) = \min\{N>1: D(T,N) > 0\},\$$

where

(2.2)
$$D(T,N) = \lambda T(av-1) - a\lambda K - (\lambda T)^{2} + (2N+1)[\lambda T + \phi_{1}(T,N)] - \phi_{2}(T,N).$$

Proof. It is easily seen that

(2.3)
$$\phi_1(T,N+1) = \phi_1(T,N) + F(N)$$

$$\phi_2(T,N+1) = \phi_2(T,N) + (2N+1)F(N),$$

where F is the Poisson distribution with mean λT . Using these expressions and (1.3) - (1.5), one can show that

 $C(T,N+1) - C(T,N) = D(T,N)(h/2)F(N)/[(\lambda T + \phi_1(T,N))(\lambda T + \phi_1(T,N+1))].$ The terms following D(T,N) are positive, and so C(T,N) will have a unique minimum over N at the value (2.1) if D(T,N) is strictly increasing in N. But this is true since one can show that

(2.4)
$$D(T,N+1) - D(T,N) = 2[\lambda T + \phi_1(T,N)] > 0.$$

Computation of Optimal N(T) Policies. The optimum N(T) in (2.1) can be obtained by computing D(T,N) recursively by the following formulas based on (2.3) and (2.4):

$$\phi_1(T,N) = \phi_1(T,N-1) + F(N-1)$$

$$D(T,N) = D(T,N-1) + 2(\lambda T + \phi_1(T,N-1)), \quad N > 2$$
 where $\phi_1(T,1) = e^{-\lambda T}$.

Our computations show that N(T), as a function of T, is nonincreasing and then nondecreasing. This was as anticipated: For T near zero, N(T) is moderate since it is the major control parameter; as T grows, N(T) can be reduced, but it eventually tends to ∞ .

Remark. Recall that the (T,N)-policy with T=0 is the N-policy. In this case, $\phi_1(0,N) = N$, $\phi_2(0,N) = N^2$ and $D(0,N) = N^2 + N - a\lambda K$; and so the optimal N(0) is the smallest integer greater than $(1/4+a\lambda K)^{1/2} - 1/2$. This is consistent with Heyman (1968).

Now consider the problem of finding an optimal (T,N)-policy. This problem can be expressed, with Theorem 2.1 in mind, as

(2.6)
$$\min C(T,N) = \min \min C(T,N) = \min C(T,N(T)).$$
T,N T N T

If the function C(T) = C(T, N(T)) were to have a unique minimum, say at

T*, then it would follow from (2.6) that (T*,N(T*)) is the unique optimal (T,N)-policy. Because the function (T) is rather intractable, we were not able to prove that it has a unique minimum. However, extensive computations showed that (T) does indeed have a unique minimum; we enumerated hundreds of functions and each one had a unique minimum.

Computation of Optimal (T,N)-Policies. From the preceding comments, it follows that an optimal (T,N)-policy can be computed as follows.

Consider a grid of T-values T_1 , T_2 ,... as fine as desired. Using (2.1) and (2.5), compute N(T) and C(T) = C(T,N(T)) for the successive T-values T_1,T_2,\ldots until the time $T^* = \min\{T_k\colon C(T_k) < C(T_{k+1})\}$. The resulting (T*,N(T*)) is the optimal (T,N)-policy. (Alternatively, one may find the T* that minimizes C(T) over T in $\{T_1,T_2,\ldots\}$ by a Fibonacci or Golden Section Search Procedure, where N(T) and C(T) are computed at each stage by (2.1) and (2.5). However, the saving of computation time by this approach is negligible.)

This procedure is very easy to implement. Examples of optimal (T,N)-policies computed by it are shown in Table 1. For these computations, we set $\lambda=1$ (which is equivalent to λ being the time unit), and we set $a=2(1-\rho)/h=1$ (which is equivalent to a^{-1} being the monetary unit).

The average cost associated with an optimal (T,N)-policy is $C(T^*,N(T^*)) = hC^*/2 + r\rho + h\lambda(1-\rho)EW,$

where C* denotes the expression in braces in (1.5), which is the only term relevant to the optimization (the other terms do not depend on (T,N)). Some of the values of C* associated with Table 1, for v=30, are as follows:

These C* values are rounded to the nearest integer. The corresponding values of C* for v below 30 are not more than one unit below these values for v=30. The C* is obviously increasing in v and K. Note that the optimal policies do not depend on the cost r of running the server or on the variance σ^2 of the service time.

Remark. If there is no cost for inspecting the queue (v=0), then it is optimal to continually inspect the queue and have no idle time (T*=0). This intuitively obvious result follows since one can show that $\frac{\partial C(T,N)}{\partial T} = 0$ when T=0.

It is of interest to know whether the optimal policy $(T^*, N(T^*))$ is nonincreasing or nondecreasing in a particular input parameter. For example, Table 1 shows that $T^* + in v$, but $N(T^*) + in v$. Here is a formal result in this regard.

Theorem 2.2.

- (i) T* is strictly \uparrow in each of the parameters K, v and μ .
- (ii) $N(T^*) + in K and + in v$.
- (iii) N(T*) † in μ for $\mu \leq \mu_0 = \inf\{\mu\colon T^* > v/\lambda^2 K\}$, and N(T*) † in μ for $\mu > \mu_0$.

<u>Proof.</u> These properties are based on the following result. Consider an optimization problem, like ours, of the form

where S is a subset of the line or plane and v>0 is a parameter of interest. Suppose f(x,v) has a minimum over $x \in S$ at the point $x^*(v)$; when there are several minima we assume there is a smallest one and call

it $x^*(v)$. That is, we assume the following minimum exists

$$x^*(v) = \min\{x: f(x,v) = \min_{x'} f(x',v)\}.$$

It is known (see for instance [8]) that $x^*(v) + or + in v$ according to whether $\frac{\partial f}{\partial v}(x,v) + or + in x$; moreover $x^*(v)$ is strictly monotone when $\frac{\partial f}{\partial v}(x,v)$ is.

First consider T^* and $N(T^*)$ as functions of μ . One can show that

$$\frac{\partial^{2}C(T,N)}{\partial T \partial \mu} = -v \lambda^{3} \mu^{-2} [F(N-1) + \phi_{1}(T,N)(1 - F(N-1))]$$

$$/[\lambda T + \phi_{1}(T,N)]^{2} < 0.$$

Thus, it follows by the preceding comments that T* is strictly \uparrow in μ (as asserted in (i)). Similarly,

$$\frac{\partial C(T,N+1)}{\partial \mu} - \frac{\partial C(T,N)}{\partial \mu} = \lambda \mu^{-2} F(N) (KT \lambda^{2} - v)$$

$$/[\lambda T + \phi_{1}(T,N+1)][\lambda T + \phi_{1}(T,N)].$$

This expression is negative or positive according to whether T is < or $> v/K\lambda^2$. This observation and T* being strictly increasing in μ proves assertion (iii). Assertion (ii) and the rest of (i) follow by similar arguments.

Remarks. Our model assumes that after each idle period, even when there are no customers waiting, the server is committed to an inspection-service period. A variation is that when a server completes an idle period and finds no customers waiting, then it takes another idle period. The results above also apply to this setting: just replace $\phi_1(T,N)$ and $\phi_2(T,N)$ by $\phi_1(T,N) - Ne^{-\lambda T}$ and $\phi_2(T,N) - N^2e^{-\lambda T}$, respectively. Note that the (0,N)-policy in this setting is not the N-policy, whereas in our model it is.

References

- [1] Balachandran, K.R. (1973). Control policies for a single server system. Management Science 19, 1013-1018.
- [2] Bell, C.E. (1971). Characterization and computation of optimal policies for operating an M/G/l queueing system with removable server. Operations Res. 19, 208-218.
- [3] Heyman, D.P. (1968). Optimal operating policies for M/G/1 queueing systems. Operations Res. 16, 362-382.
- [4] Heyman, D.P. and M.J. Sobel (1982). Stochastic Models in Operations Research Vol. I. McGraw Hill.
- [5] Heyman, D.P. (1977). The T-policy for the M/G/1 queue. Management Science 23, 775-778.
- [6] Kimura, T. (1982). Further results on the N-policy for the M/G/l queue. Technical report. Department of Information Sciences, Tokyo Institute of Technology.
- [7] Levy, Y. and U. Yechiali (1975). Utilization of idle time in an M/G/1 queueing system. Management Science 22, 202-211.
- [8] Lu, F.V. and R.F. Serfozo (1984). M/M/l queueing decision processes with monotone hysteretic optimal policies. Operations Res. 32, 1116-1132.
- [9] Shanthikumar, J.G. (1981). Optimal control of an M/G/l priority queue via N-control. American Journal of Mathematical and Management Sciences 1, 191-212.
- [10] Sobel, M.J. (1969). Optimal average cost policies for a queue with start-up and shut-down costs. Operations Res. 17, 145-162.
- [11] Talman, A. (1979). A simple proof of the optimality of the best N-policy in the M/G/l queueing control problem with removable server. Statist. Neerlandica 33, 143-150.
- [12] Tijms, H.C. (1976). Optimal control of the workload in an M/G/l queueing system with removable server. Math. Operationsforsch. Statist. 6, 933-943.
- [13] Yadin, M and P. Naor (1963). Queueing systems with a removable service station. Operations Research Quarterly 14, 393-405.

Table 1 $\label{eq:continuous} \mbox{OPTIMAL (T,N)-POLICIES}$ The Table entries are T* N(T*).

K		0	5		10		15		20		25		30	_	_
100	9	0	9.5	7	9.9	4	10.0	2	10.0	1	10.0	1	10.0	1	
120	10	0	10.4	8	10.9	5	11.0	3	11.0	1	11.0	1	11.0	1	l
140	11	0	11.2	9	11.7	6	11.8	4	11.8	1	11.8	1	11.8	1	l
160	12	0	12.0	10	12.5	7	12.6	5	12.6	2	12.6	1	12.6	1	l
180	12	0	12.7	10	13.3	8	13.4	5	13.4	3	13.4	1	13.4	1	ĺ
200	13	0	13.8	11	14.0	9	14.1	6	14.1	4	14.1	1	14.1	1	١
220	14	0	14.0	12	14.7	9	14.8	7	14.8	4	14.8	2	14.8	1	
240	14	0	14.6	12	15.3	10	15.5	7	15.5	5	15.5	2	15.5	1	ĺ
260	15	0	15.2	13	15.9	11	16.1	8	16.1	6	16.1	3	16.1	1	l
280	16	0	15.8	14	16.5	11	16.7	9	16.7	6	16.7	4	16.7	1	
300	16	0	16.3	14	17.1	12	17.3	9	17.3	7	17.3	4	17.3	2	l
320	17	0	16.8	15	17.6	12	17.8	10	17.9	7	17.9	5	17.9	2	
340	17	0	17.3	15	18.2	13	18.4	10	18.4	8	18.4	5	18.4	3	l
360	18	0	17.8	16	18.7	13	18.9	11	19.0	8	19.0	6	19.0	3	l
380	18	0	18.3	16	19.2	14	19.4	11	19.5	9	19.5	6	19.5	4	
400	19	0	18.8	17	19.7	14	19.9	12	20.0	9	20.0	7	20.0	4	l
420	19	0	19.2	17	20.2	15	20.4	12	20.5	10	20.5	7	20.5	5	
440	20	0	19.7	18	20.6	15	20.9	13	21.0	10	21.0	8	21.0	5	
460	20	0	20 - 1	18	21.1	16	21.4	13	21.4	11	21.4	8	21.4	6	
480	21	0	20.6	19	21.5	16	21.8	14	21.9	11	21.9	9	21.9	6	
500	21	0	21.0	19	22.0	17	22.3	14	22.3	12	22.4	9	22.4	7	
	i		Ī		1				ľ		ł		!		Ĺ

5/

in intellection